

Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models

Year: 2024 | Citations: 455 | Authors: Yixin Liu, Kai Zhang, Yuan Li

Abstract

Sora is a text-to-video generative AI model, released by OpenAI in February 2024. The model is trained to generate videos of realistic or imaginative scenes from text instructions and show potential in simulating the physical world. Based on public technical reports and reverse engineering, this paper presents a comprehensive review of the model's background, related technologies, applications, remaining challenges, and future directions of text-to-video AI models. We first trace Sora's development and investigate the underlying technologies used to build this "world simulator". Then, we describe in detail the applications and potential impact of Sora in multiple industries ranging from film-making and education to marketing. We discuss the main challenges and limitations that need to be addressed to widely deploy Sora, such as ensuring safe and unbiased video generation. Lastly, we discuss the future development of Sora and video generation models in general, and how advancements in the field could enable new ways of human-AI interaction, boosting productivity and creativity of video generation.